Probabilistic Structural Biology: The Challenge of Determining, Describing and Predicting the Structure of a Flexible Protein

CASP-16
Punta Cuna, Dominican Republic
December 4th, 2024

Terry Oas, Departments of Biochemistry and Chemistry
Bruce Donald, Departments of Computer Science and Biochemistry
Duke University
Durham, NC

Probabilistic Models of Biomolecular Structure

- The goal of experimental structural biology is to make useful models of bimolecular atomic structure based on *experimental observations*.

- The goal of CASP is to aid in the development of a generalized predictive model based on *sequence*.

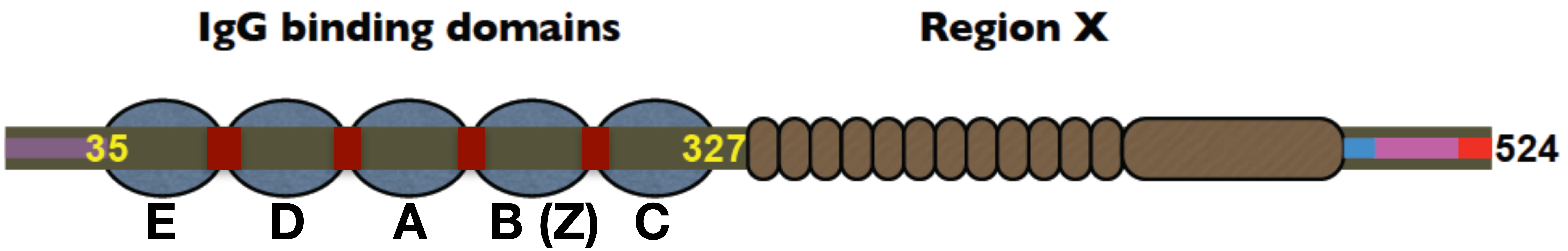
- Conventional structural biology can't be used to build a model of a flexible protein.

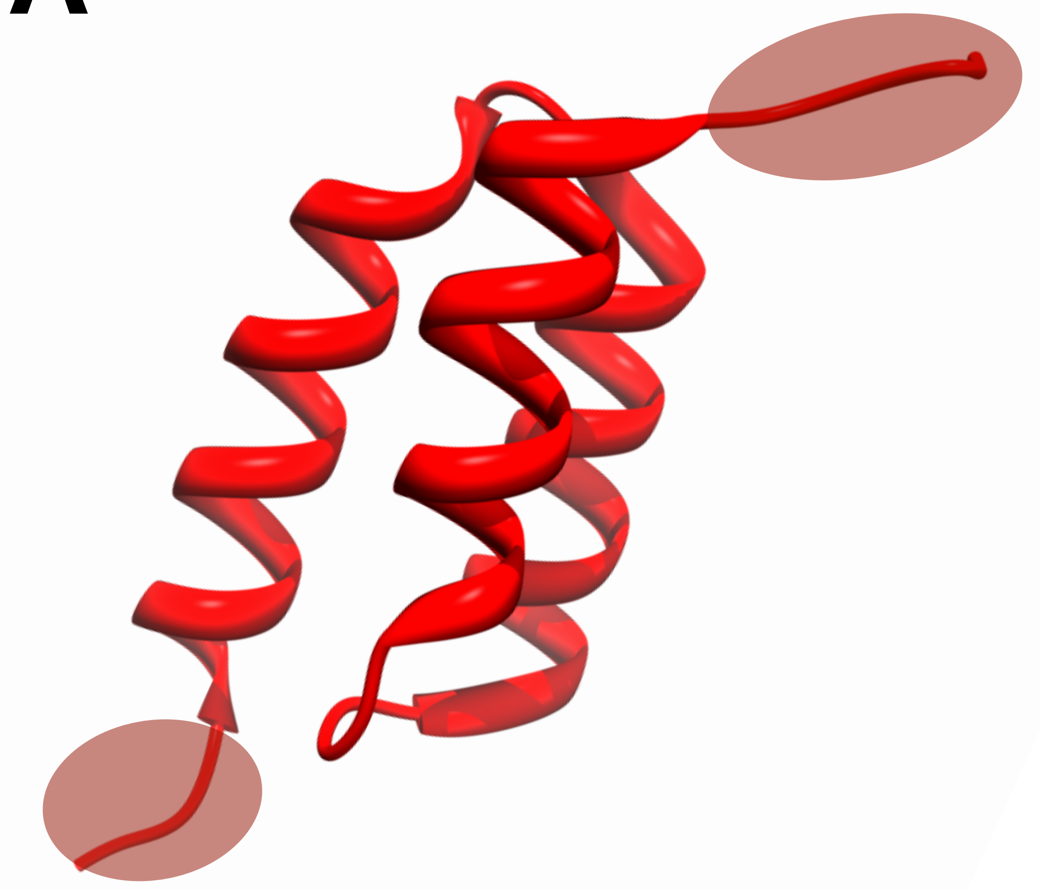
- The interdomain orientation and distance of a domain-linker-domain (D-L-D) protein are determined by a *continuous* probability field on the scale of the domains.

- Therefore, the structure of a D-L-D protein is best described with a *continuous* probability distribution.

- NMR residual dipolar couplings can provide information about the continuous distribution of interdomain orientation (CDIO.)

ZLBT-C: A D-L-D protein from Staphylococcal protein A has a flexible interdomain linker (KADNKF)





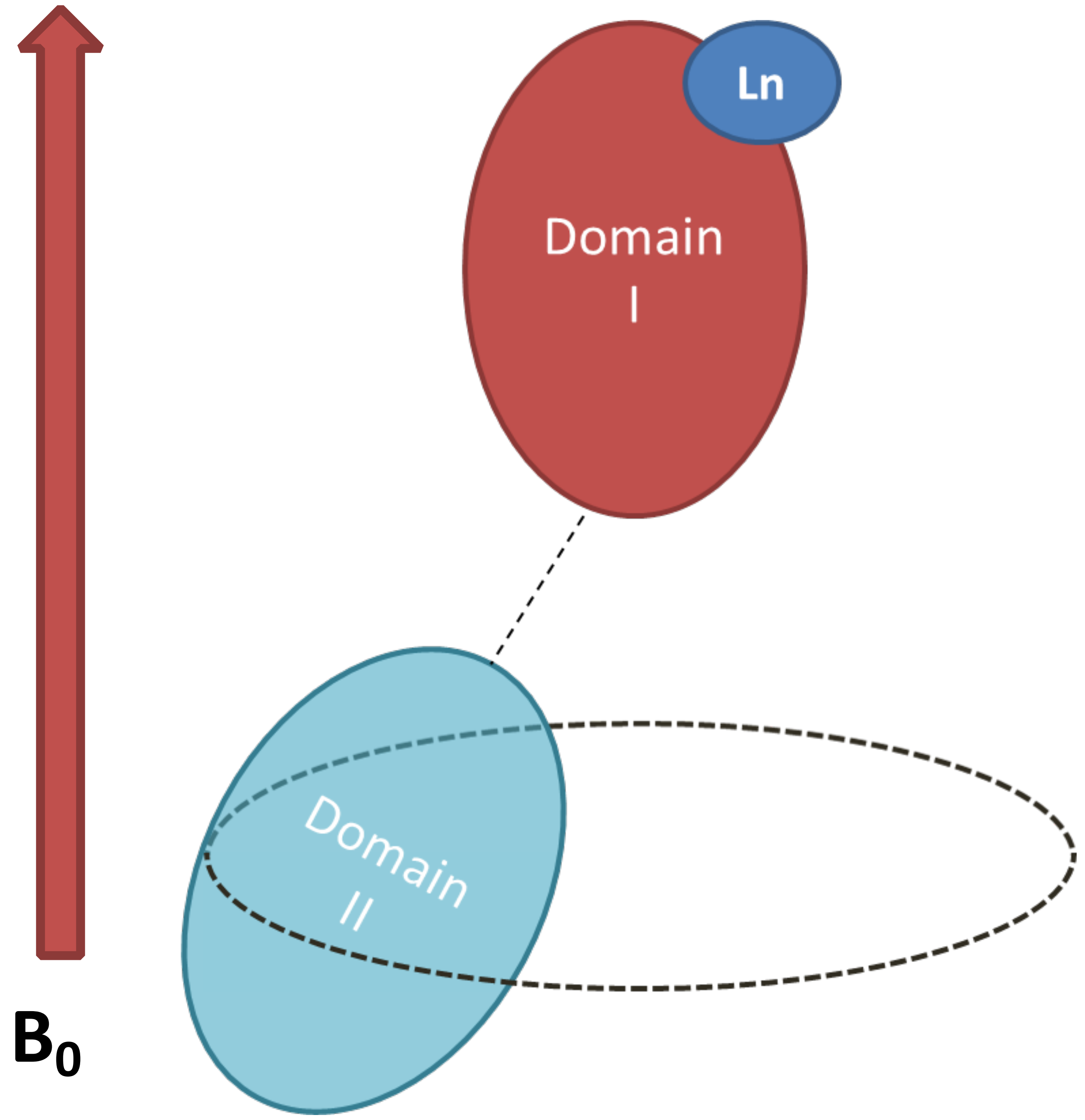
Magnetic alignment using lanthanides

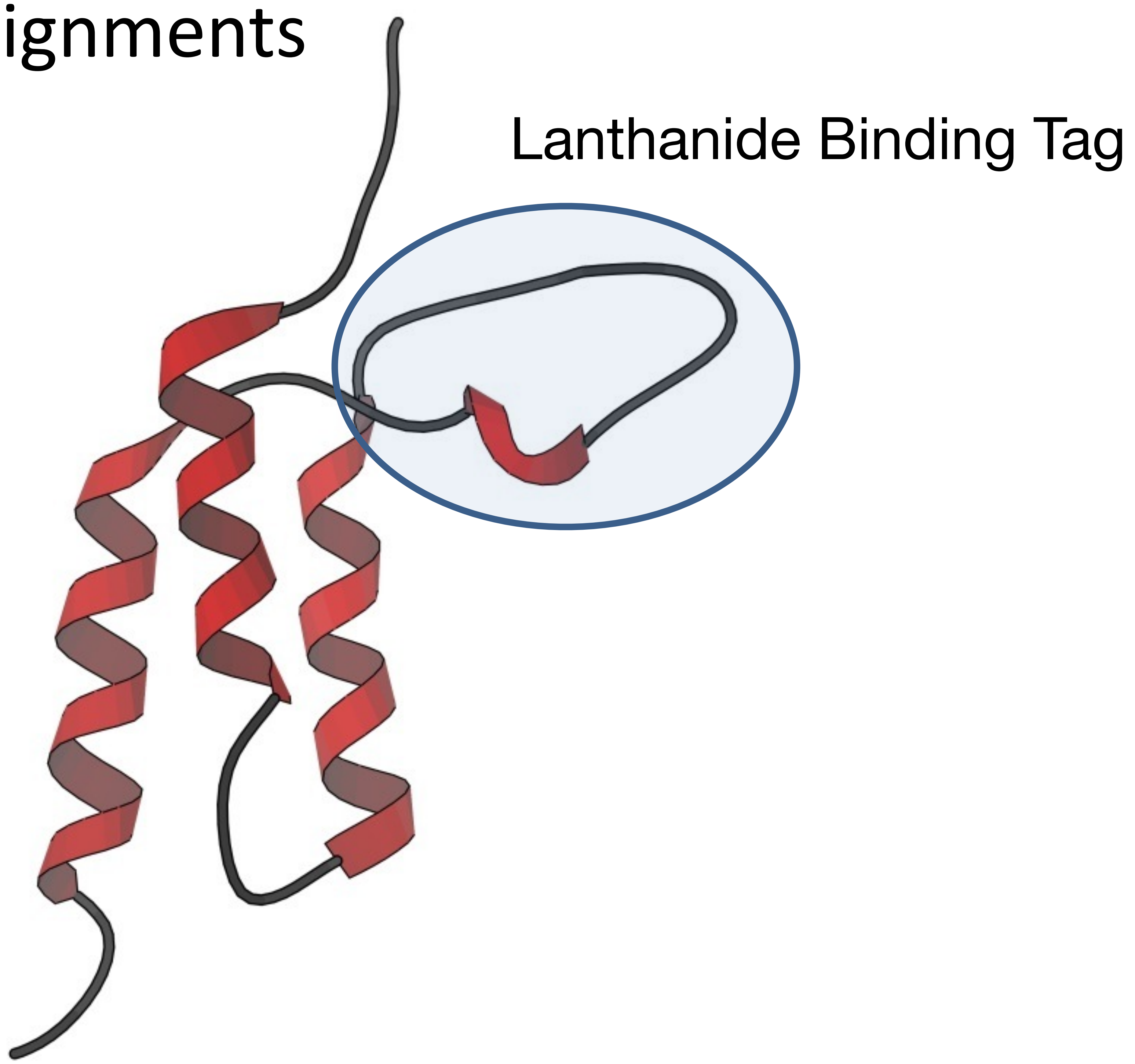
Conformation decoupling:

The alignment of domain I is independent of inter-domain conformations

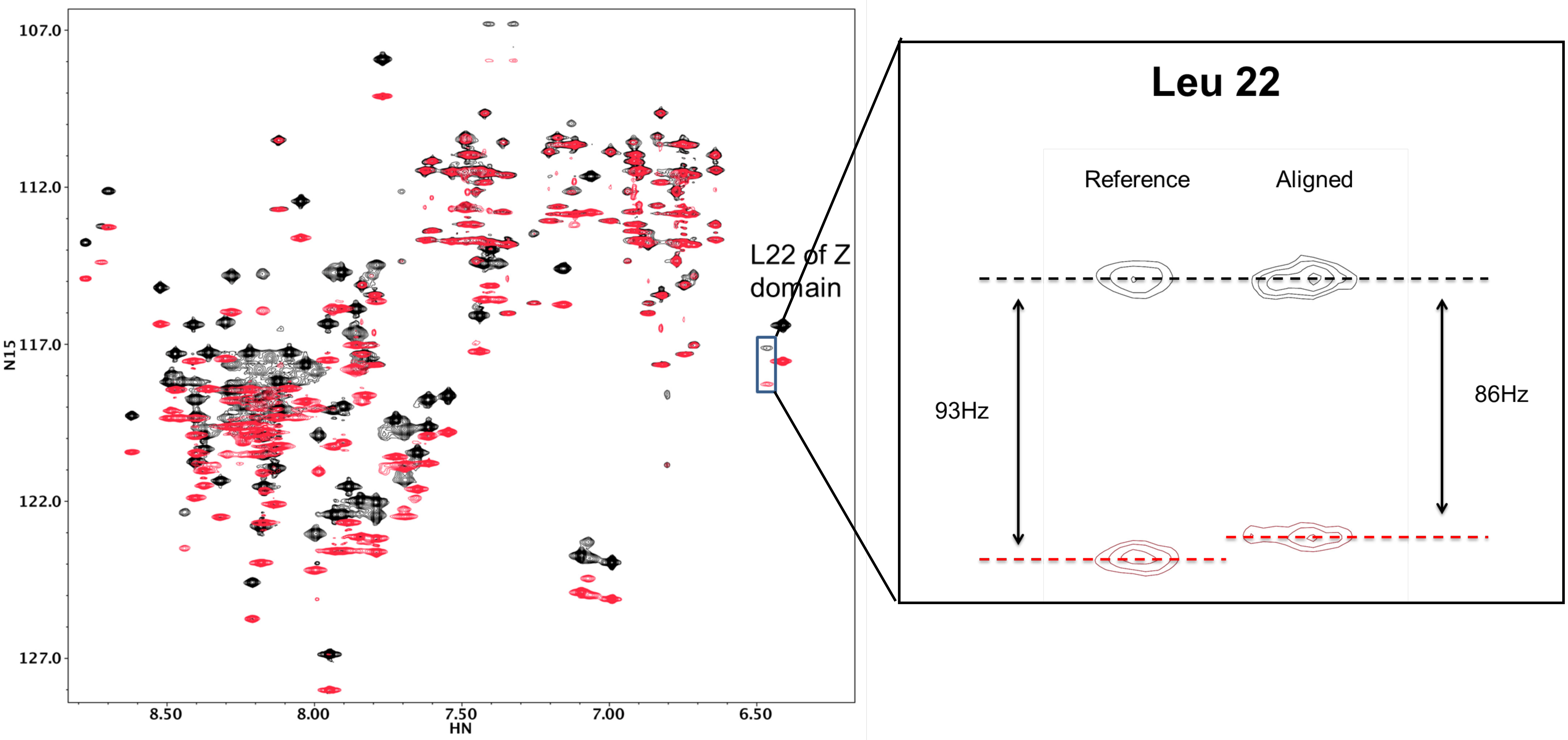
Multiple alignments:

Using different lanthanides gives different alignments

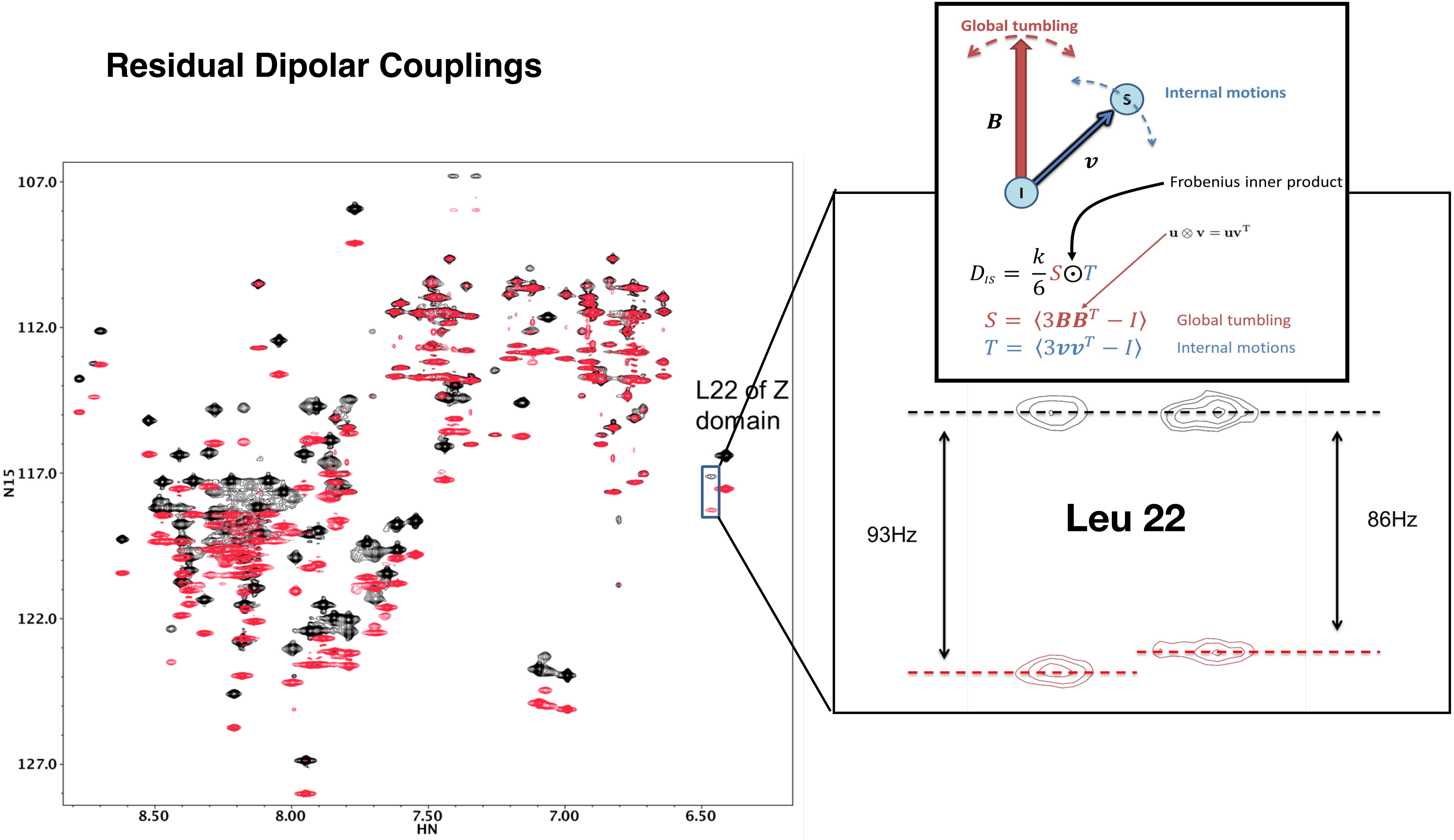




Residual Dipolar Coupling Measurement

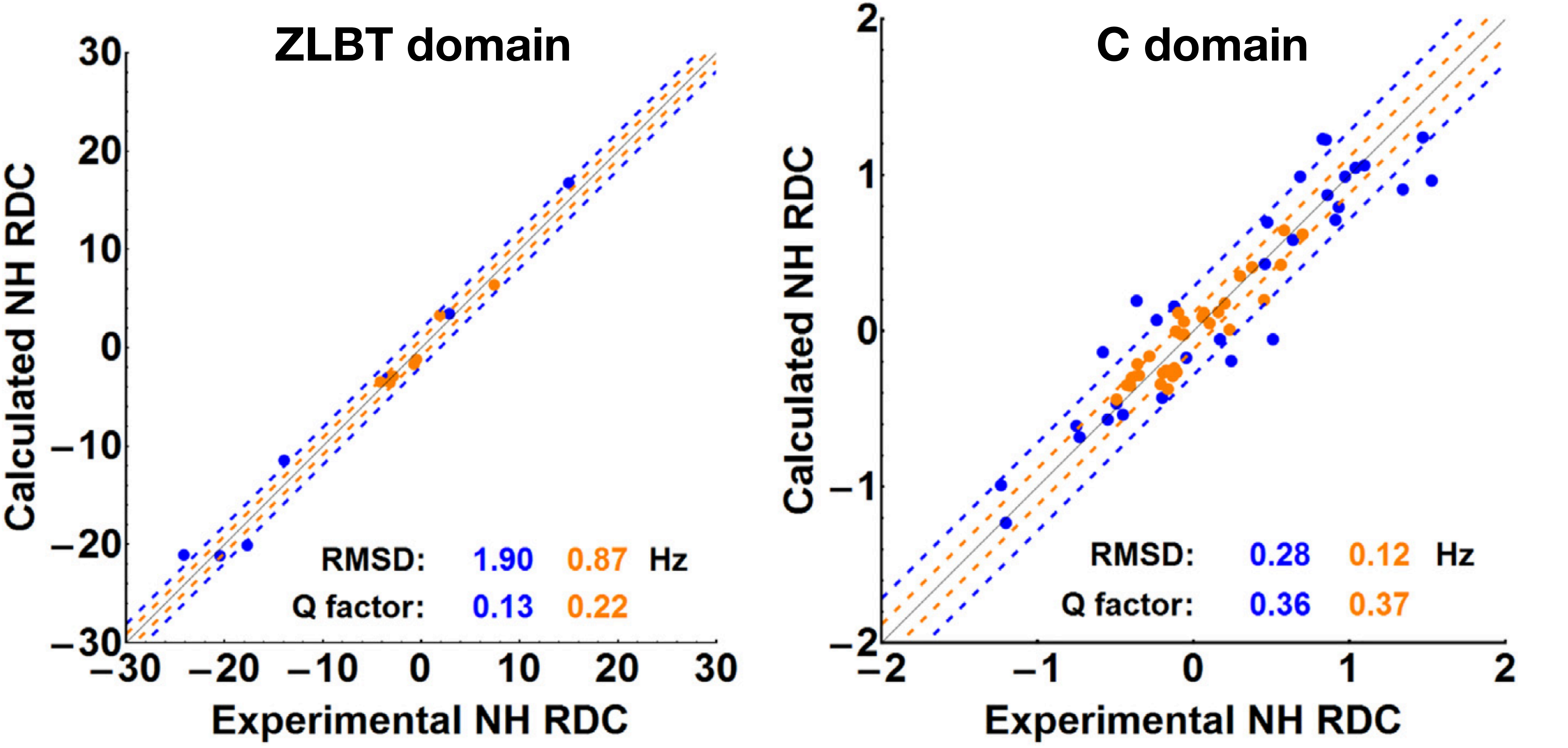


Residual Dipolar Couplings

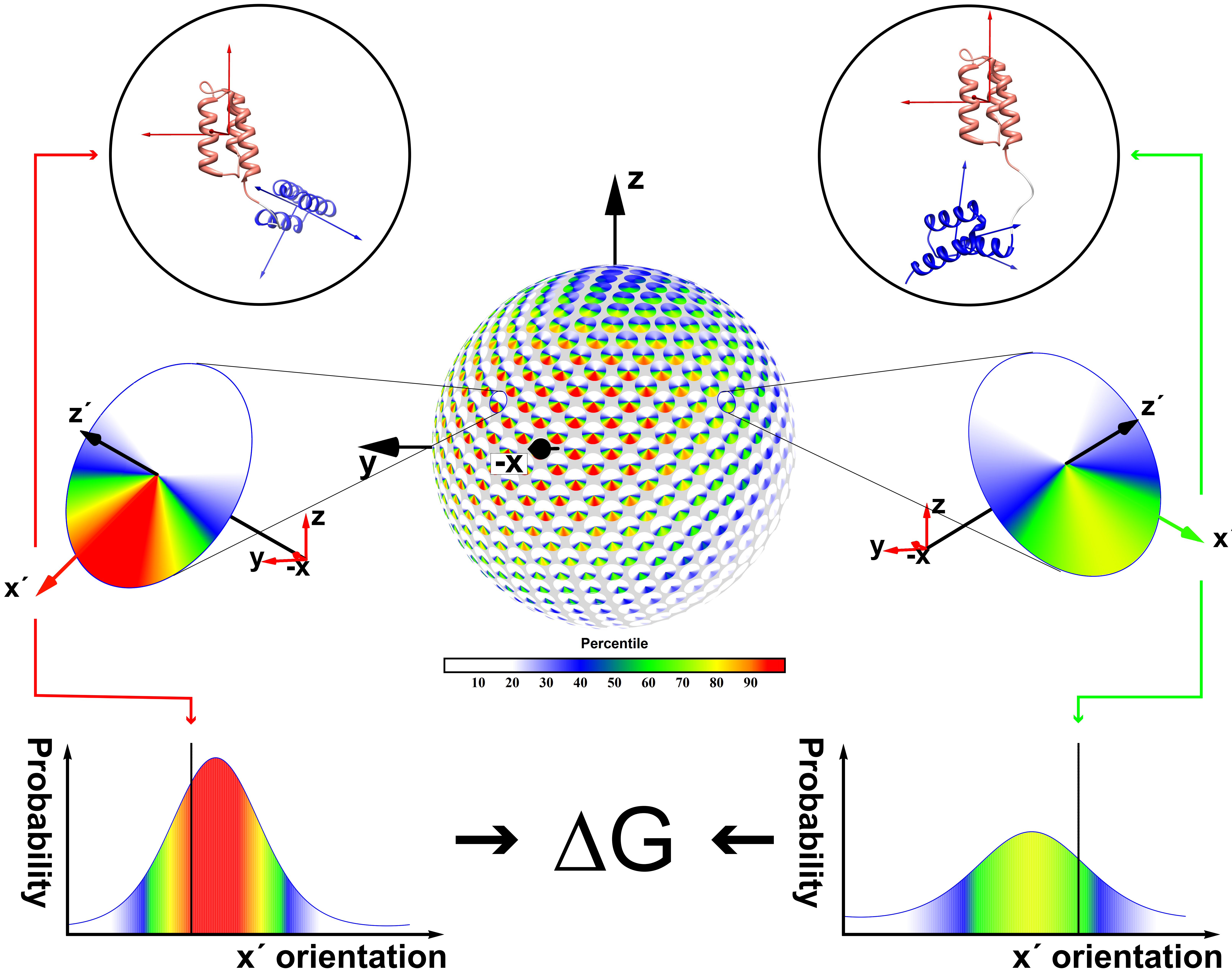


RDC values back-calculated from the computed alignment tensors agree well with the experimental values

Experimental RDC uncertainties are ±0.2 HZ

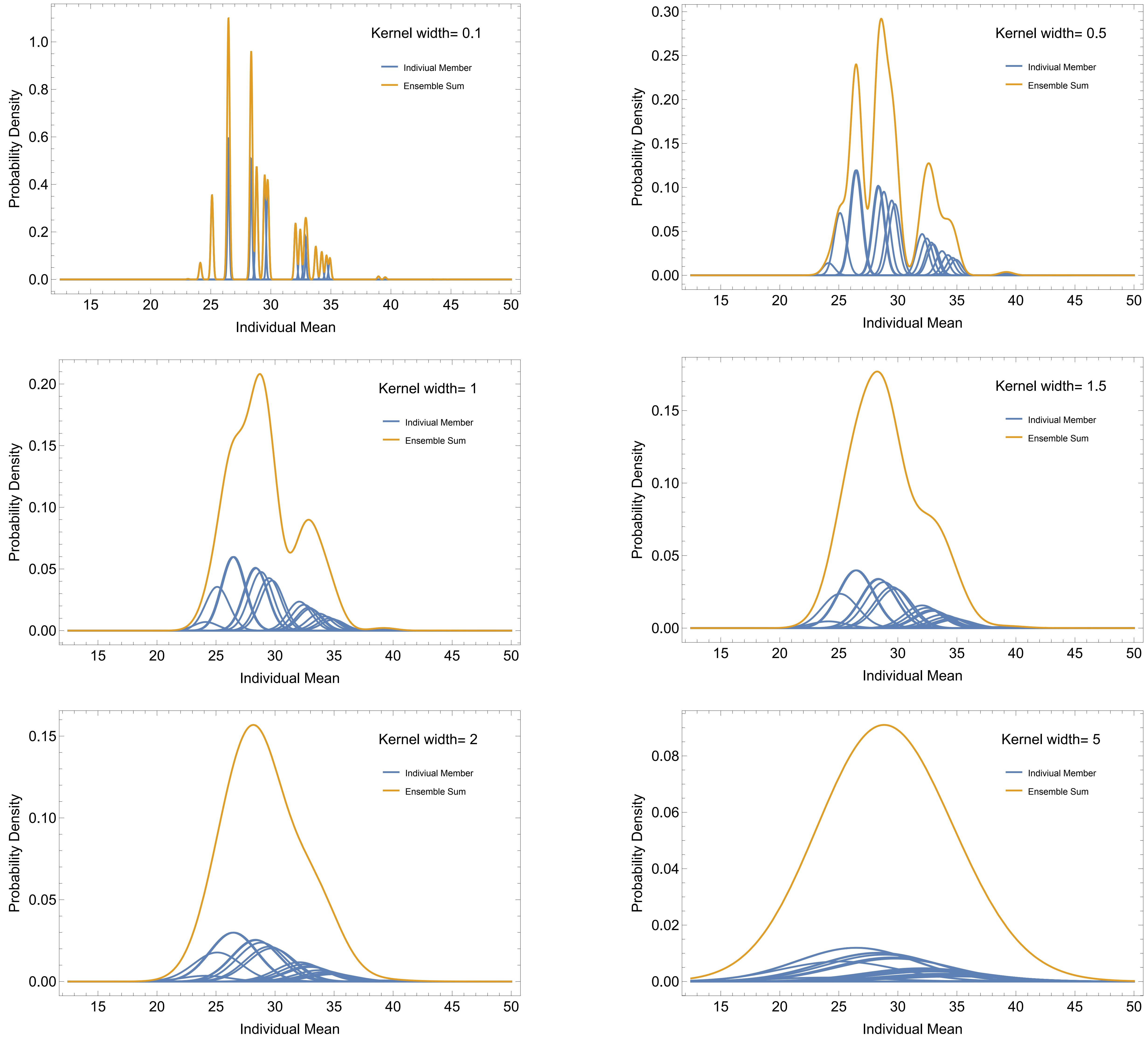


Disk-on-Sphere representation of a Continuous Distribution of Interdomain Orientation (CDIO)

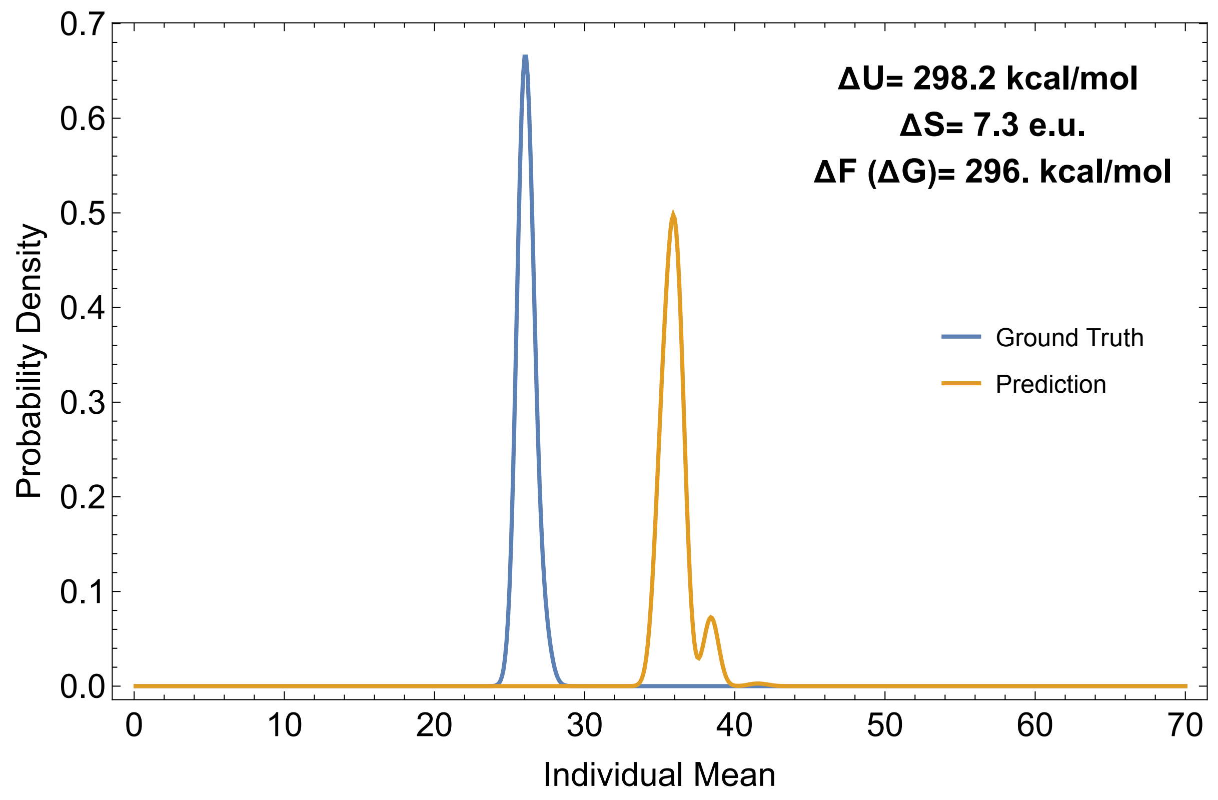


Discrete ensembles can be converted to CDIO's by representing each member by its orientation with a population-weighted kernel function.

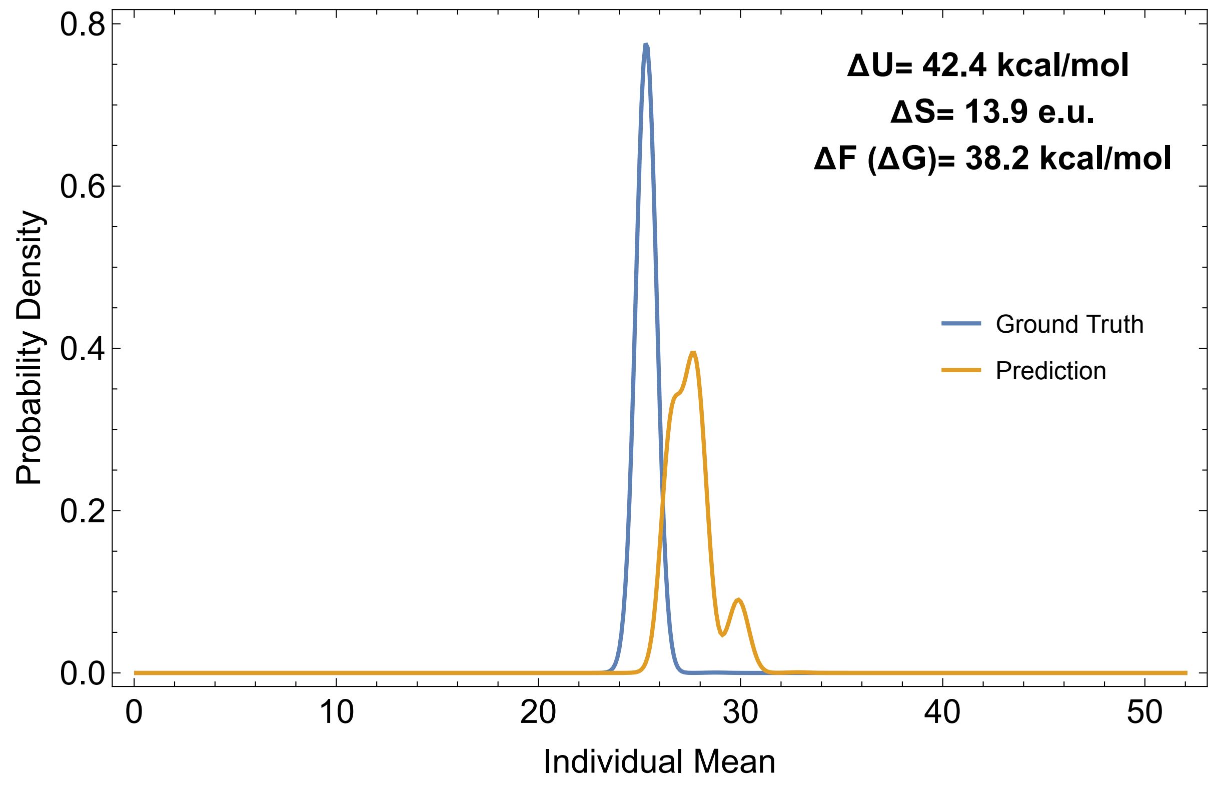
2D Example:



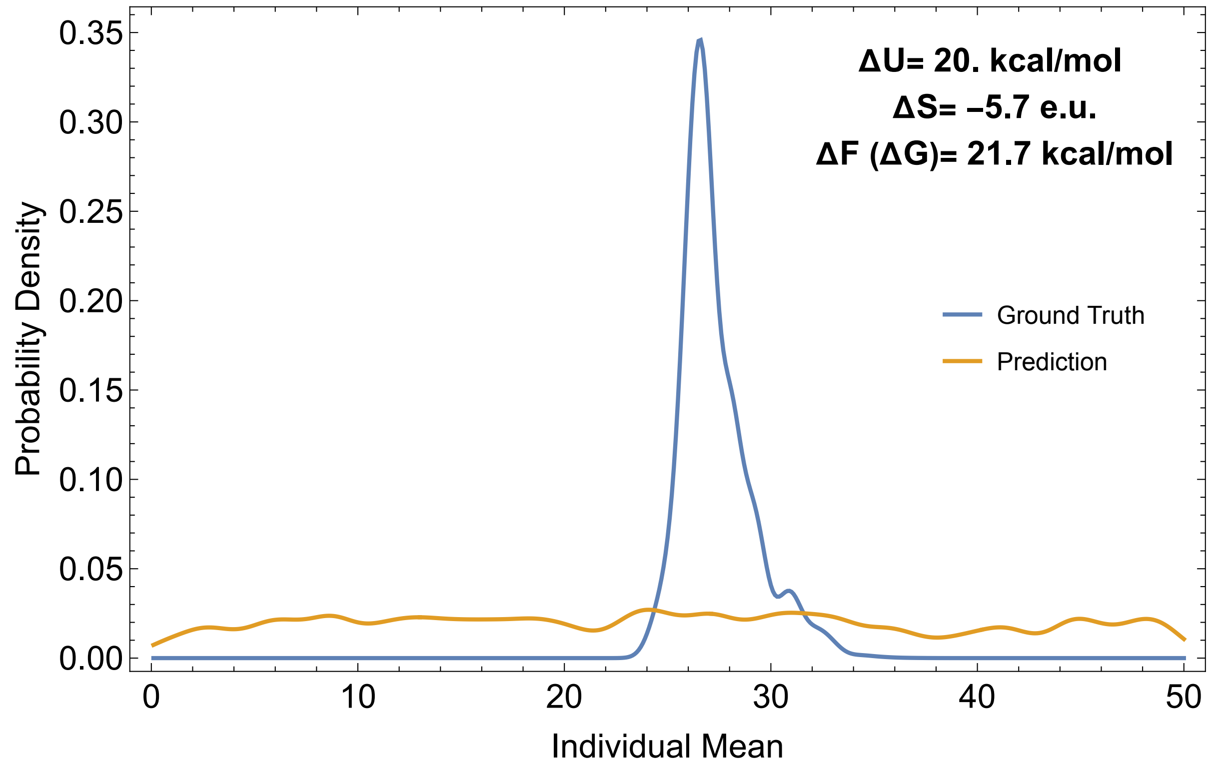
By the free energy metric, isotropic predictions always beat anisotropic ones



Anisotropic prediction with a very different mode

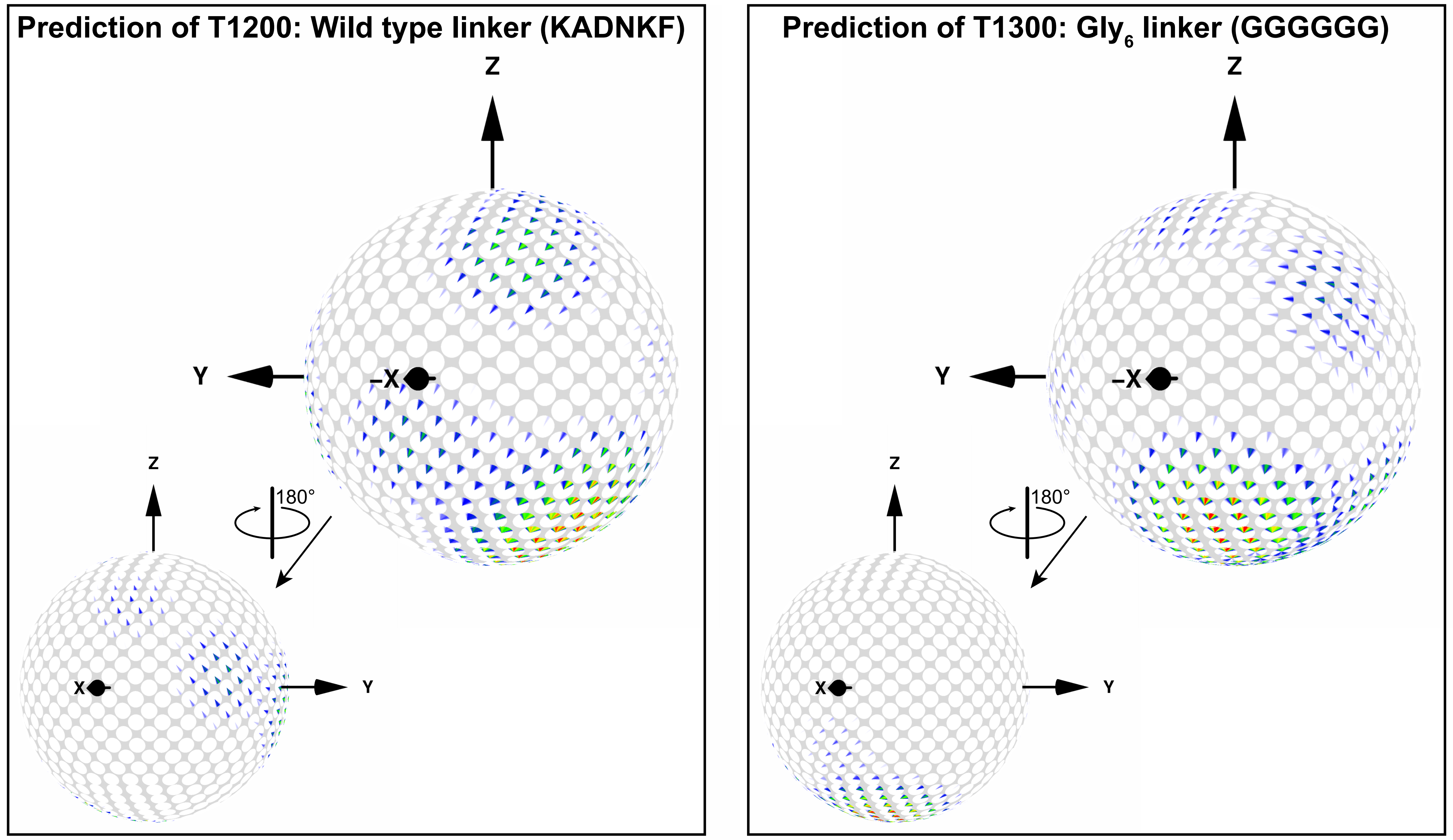


Anisotropic prediction with a similar mode



Isotropic prediction with undefined mode

Comparison Between Predicted and Observed SO(3) Probability Distributions
Group 465: Wallner



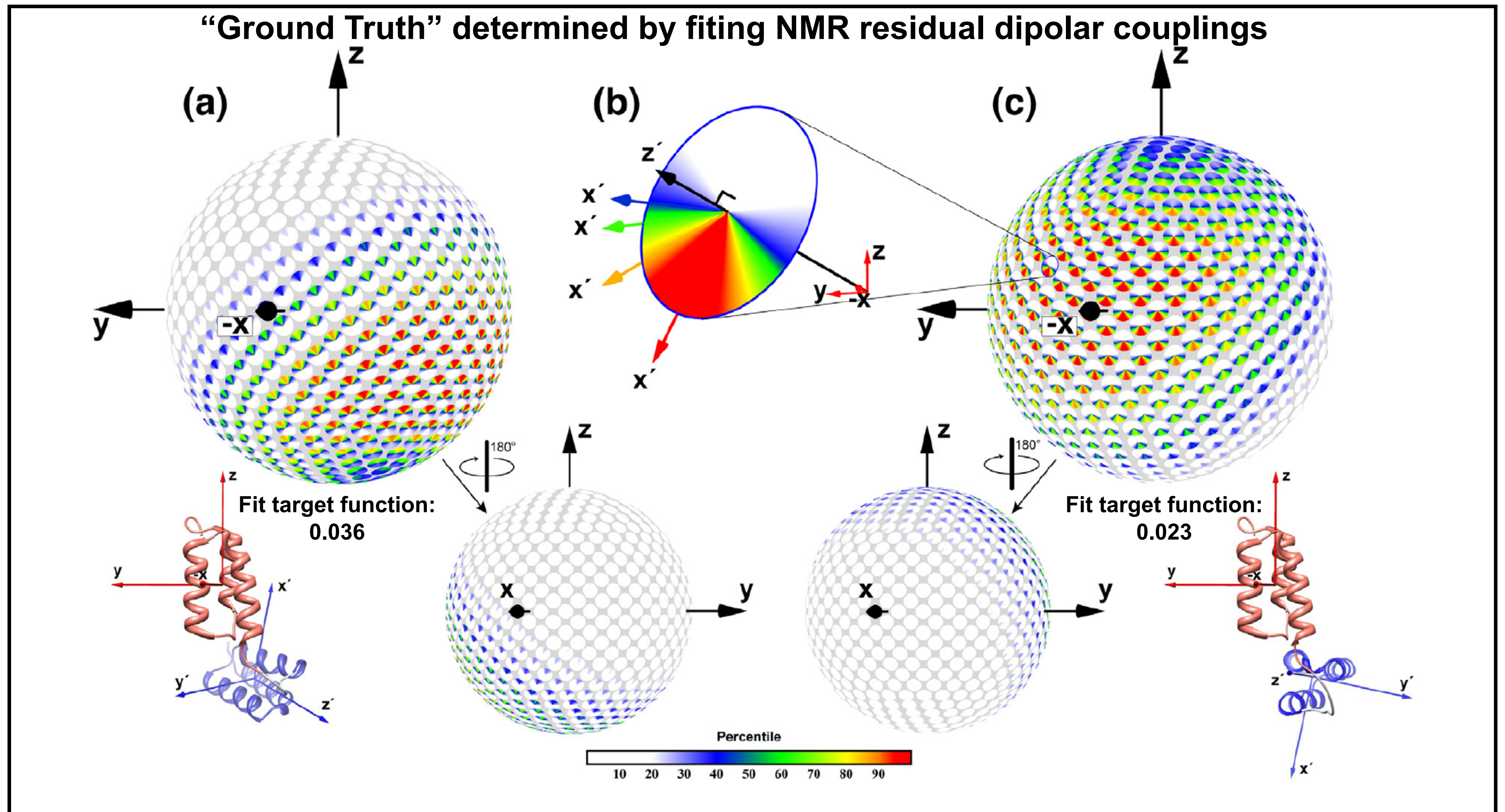
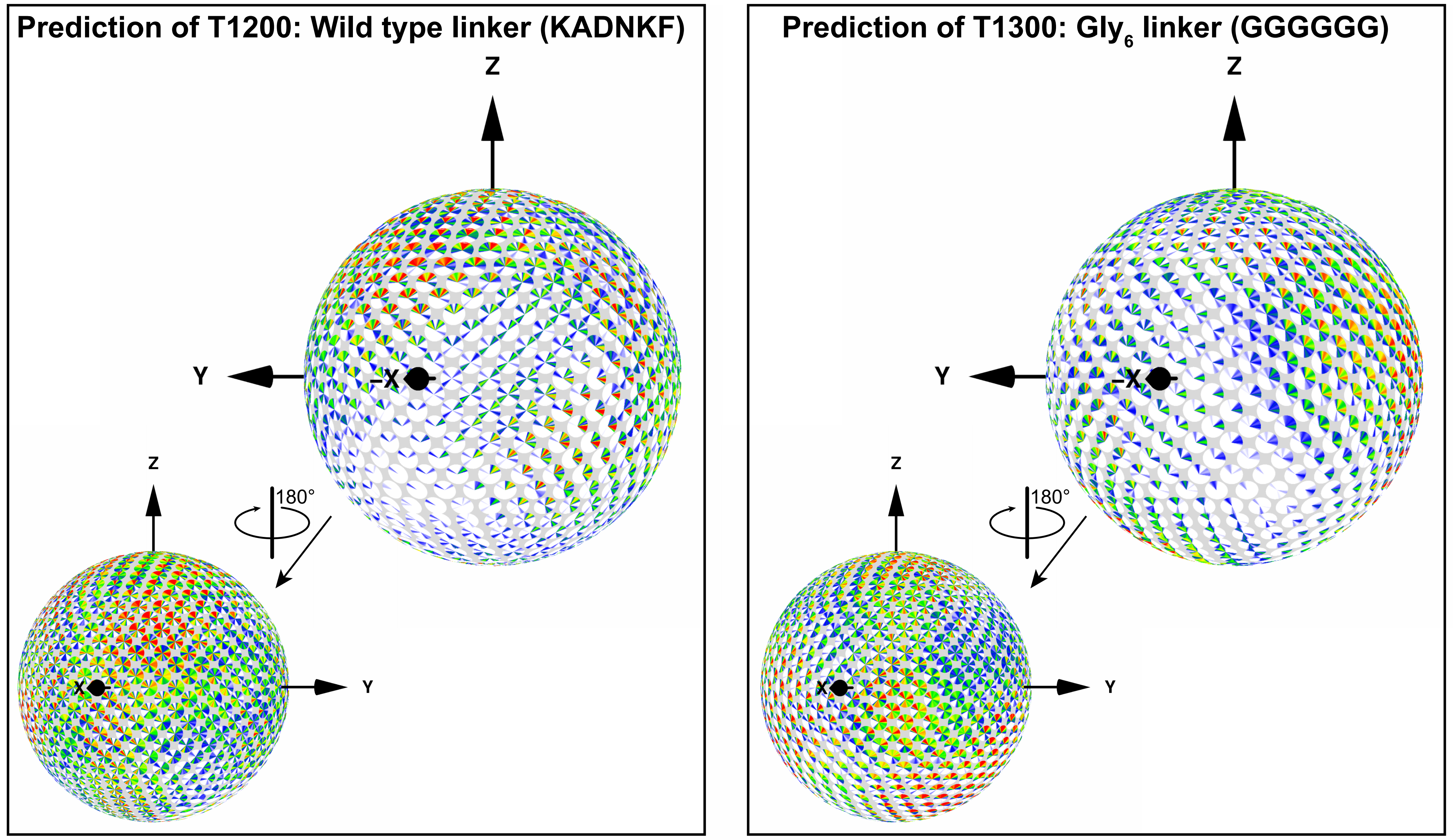


Fig. 5. CDIO models for ZLBT-C shown in DoS views. Two solutions give equivalently good fits to the data. The first solution is shown in panel a and the second solution is shown in panel c. Each panel shows the distribution from a front view (top) and a back view (bottom). Also shown are atomistic models whose interdomain orientation is the most probable in each solution. The linker conformation and interdomain distance are arbitrary. b, An example disk showing the joint probabilities of four different interdomain orientations, all with the same *z′* axis orientation but different rotations around *z′*.

Comparison Between Predicted and Observed SO(3) Probability Distributions
Group 135: Lindorff-LarsenCLVDS



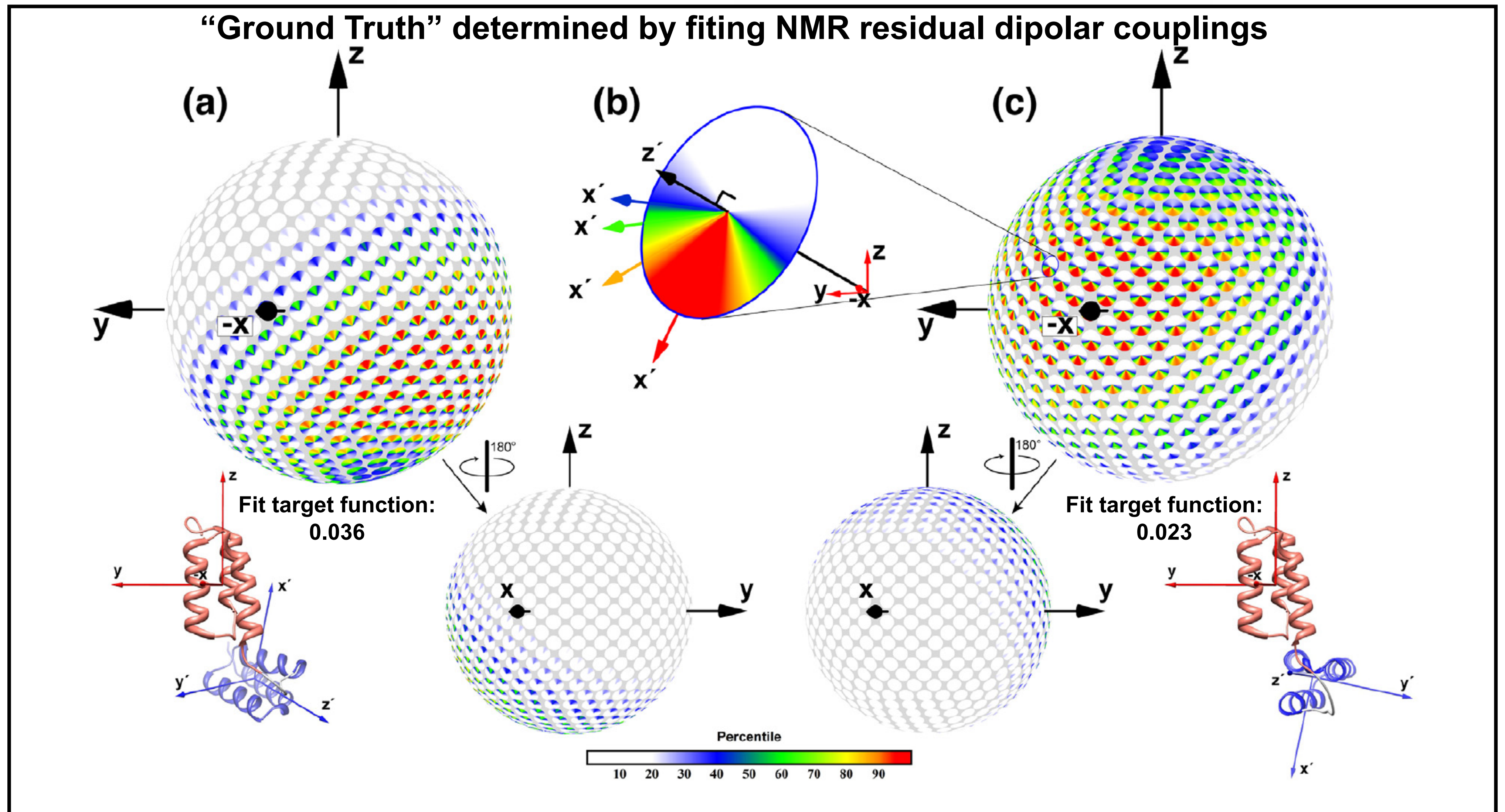


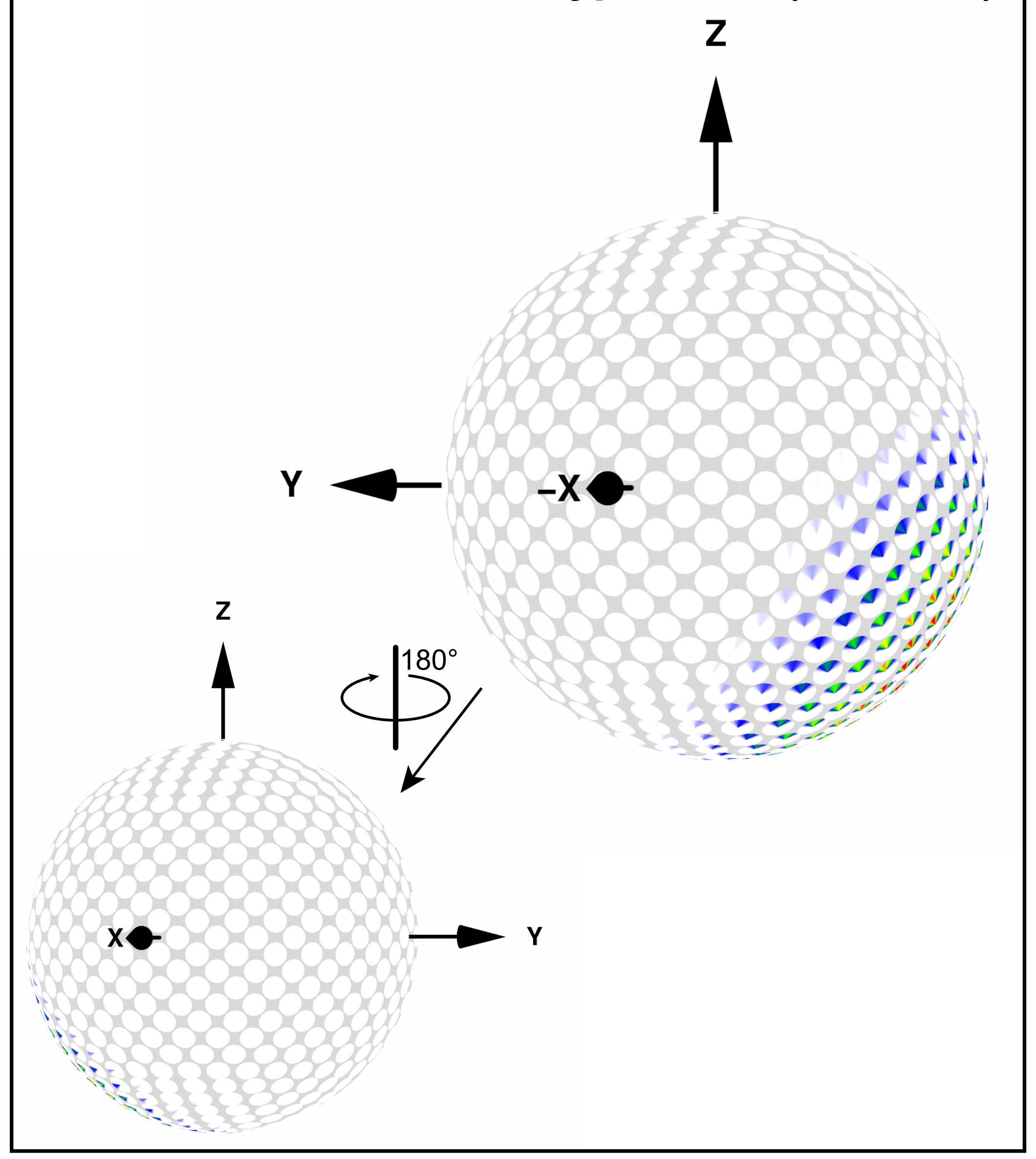
Fig. 5. CDIO models for ZLBT-C shown in DoS views. Two solutions give equivalently good fits to the data. The first solution is shown in panel a and the second solution is shown in panel c. Each panel shows the distribution from a front view (top) and a back view (bottom). Also shown are atomistic models whose interdomain orientation is the most probable in each solution. The linker conformation and interdomain distance are arbitrary. b, An example disk showing the joint probabilities of four different interdomain orientations, all with the same *z′* axis orientation but different rotations around *z′*.

Comparison Between Predicted and Observed SO(3) Probability Distributions
Group 167: OpenComplex

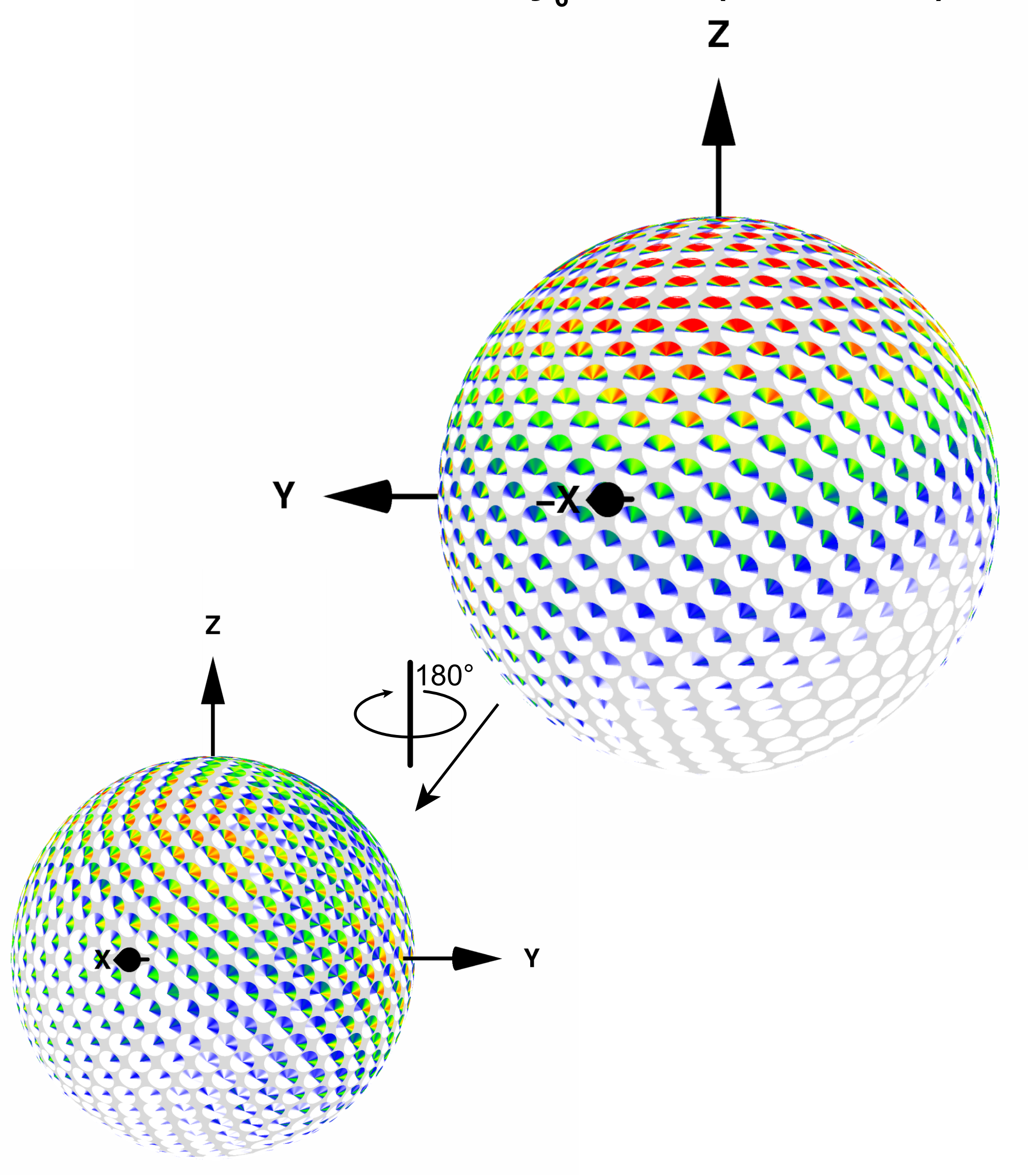
Anisotropic wild type (T1200) prediction

Qualitatively agrees with experiment

Prediction of T1200: Wild type linker (KADNKF)



Prediction of T1300: Gly_6 linker (GGGGGG)



Nearly isotropic Gly_6 (T1300) prediction

Qualitatively agrees with experiment

“Ground Truth” determined by fiting NMR residual dipolar couplings

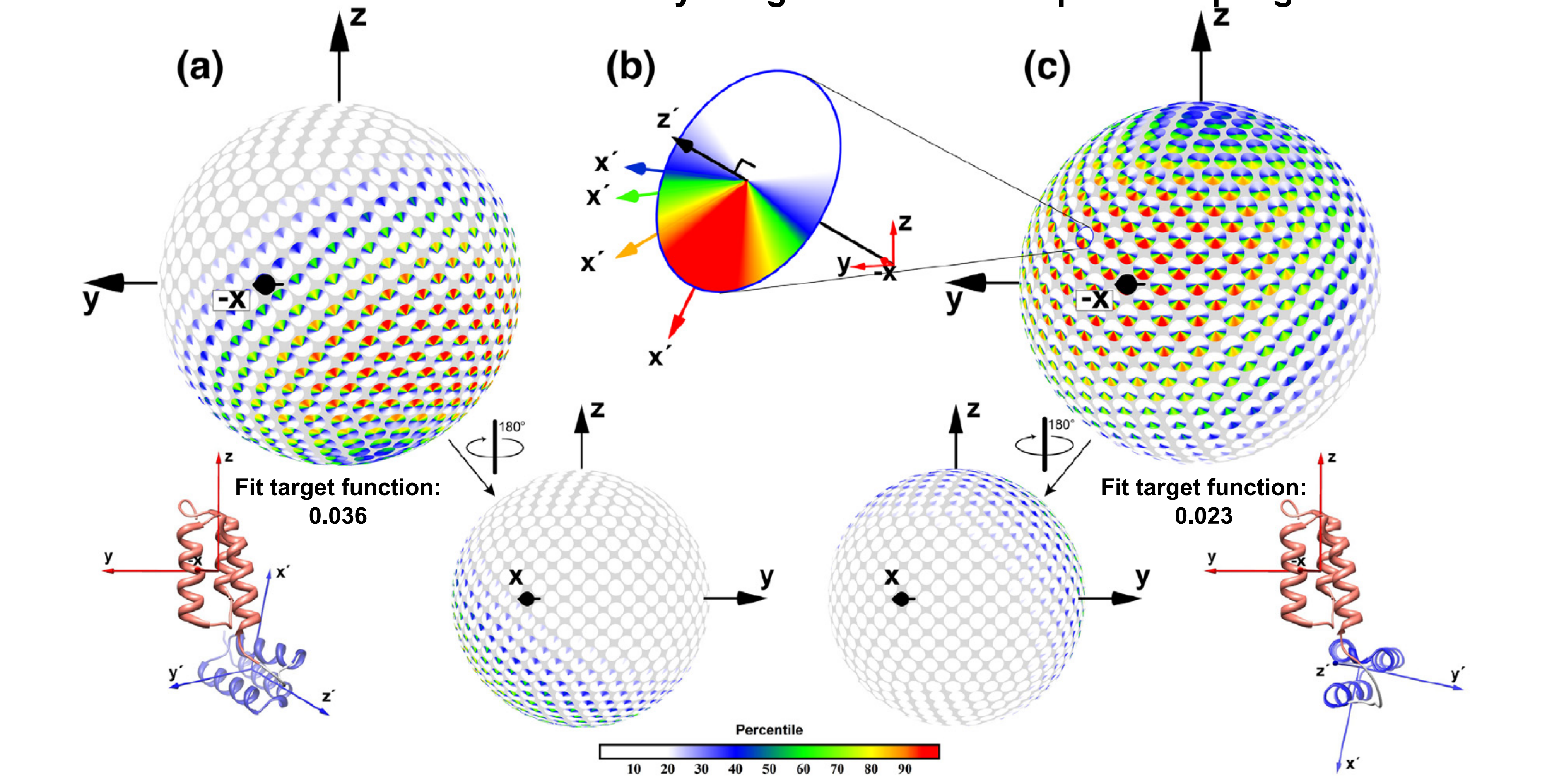


Fig. 5. CDIO models for ZLBT-C shown in DoS views. Two solutions give equivalently good fits to the data. The first solution is shown in panel a and the second solution is shown in panel c. Each panel shows the distribution from a front view (top) and a back view (bottom). Also shown are atomistic models whose interdomain orientation is the most probable in each solution. The linker conformation and interdomain distance are arbitrary. b, An example disk showing the joint probabilities of four different interdomain orientations, all with the same *z*′ axis orientation but different rotations around *z*′.

Acknowledgments

The CDIO Donald Lab Team:

- Dr. Allen McBride, Research Scientist
- Edward Cheng, Graduate Student
- Aulane Mpouli, Graduate Student

The Oas Lab:

Dr. Yang Qi, Duke Biochemistry

Dr. Lindsay Deis, Duke Biochemistry

Dr. JoAnna Capp, Duke Biochemistry

Dr. Susan Tsutakawa, LBNL

Dr. Feng Yu, LBLN

Dr. Ron Venters, Duke NMR Center

Funding:

NIGMS: R35 GM-144042 to BRD

Duke School of Medicine to TGO